Question 1. Given any unit vector $\boldsymbol{n}$ (i.e. $||\boldsymbol{n}|| = 1$), we define the hyperplane $\mathcal{H}_{\boldsymbol{n}} := \{\boldsymbol{x} : \boldsymbol{n}^\top \boldsymbol{x} = 0\}$ for which $\boldsymbol{n}$ is known as the normal vector. For any vector $\boldsymbol{x}$, we define its projection into $\mathcal{H}_{\boldsymbol{n}}$ as $\pi_{\boldsymbol{n}}(\boldsymbol{x}) = \boldsymbol{x} - (\boldsymbol{x}^\top \boldsymbol{n})\boldsymbol{n}$.

1. Given two vectors $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, take $\boldsymbol{n} = \frac{\boldsymbol{x}_2 - \boldsymbol{x}_1}{||\boldsymbol{x}_2 - \boldsymbol{x}_1||}$. Show that $\pi_{\boldsymbol{n}}(\boldsymbol{x}_1) = \pi_{\boldsymbol{n}}(\boldsymbol{x}_2)$.
2. Let $\boldsymbol{w}$ be a vector and define $y_1 := \boldsymbol{x}_1^\top \boldsymbol{w}$ and $y_2 := \boldsymbol{x}_2^\top \boldsymbol{w}$. Show that $y_1 = y_2$ if and only if $\boldsymbol{w} \in \mathcal{H}_{\boldsymbol{n}}$.

*3. Let $\boldsymbol{X}$ be a n by p matrix whose rows $\boldsymbol{X}_{i,:}$ are all distinct. Show that there exists a vector $\boldsymbol{w}$ of length p such that the scalars $(\boldsymbol{X}\boldsymbol{w})_i$ are all distinct.

Answer 1. 1.

$$\boldsymbol{n} = \frac{\boldsymbol{x}_2 - \boldsymbol{x}_1}{||\boldsymbol{x}_2 - \boldsymbol{x}_1||} \implies \boldsymbol{x}_1 = \boldsymbol{x}_2 - ||\boldsymbol{x}_2 - \boldsymbol{x}_1||\boldsymbol{n}$$

$$\begin{aligned}
\pi_{\boldsymbol{n}}(\boldsymbol{x}_1) &= \boldsymbol{x}_1 - (\boldsymbol{x}_1^\top \boldsymbol{n})\boldsymbol{n} \\
&= (\boldsymbol{x}_2 - ||\boldsymbol{x}_2 - \boldsymbol{x}_1||\boldsymbol{n}) - [(\boldsymbol{x}_2 - ||\boldsymbol{x}_2 - \boldsymbol{x}_1||\boldsymbol{n})^\top \boldsymbol{n}]\boldsymbol{n} \\
&= \boldsymbol{x}_2 - ||\boldsymbol{x}_2 - \boldsymbol{x}_1||\boldsymbol{n} - [\boldsymbol{x}_2^\top \boldsymbol{n} - ||\boldsymbol{x}_2 - \boldsymbol{x}_1||(\boldsymbol{n}^\top \boldsymbol{n})]\boldsymbol{n} \\
&= \boldsymbol{x}_2 \underline{- ||\boldsymbol{x}_2 - \boldsymbol{x}_1||\boldsymbol{n}} - (\boldsymbol{x}_2^\top \boldsymbol{n})\boldsymbol{n} \underline{+ ||\boldsymbol{x}_2 - \boldsymbol{x}_1||\boldsymbol{n}} \qquad (\because \boldsymbol{n}^\top \boldsymbol{n} = 1) \\
&= \boldsymbol{x}_2 - (\boldsymbol{x}_2^\top \boldsymbol{n})\boldsymbol{n} \\
&= \pi_{\boldsymbol{n}}(\boldsymbol{x}_2)
\end{aligned}$$

$\therefore \pi_{\boldsymbol{n}}(\boldsymbol{x}_1) = \pi_{\boldsymbol{n}}(\boldsymbol{x}_2)$

2. $y_2 - y_1 = \boldsymbol{x}_2^\top \boldsymbol{w} - \boldsymbol{x}_1^\top \boldsymbol{w} = (\boldsymbol{x}_2^\top - \boldsymbol{x}_1^\top)\boldsymbol{w} = (\boldsymbol{x}_2 - \boldsymbol{x}_1)^\top \boldsymbol{w} = ||\boldsymbol{x}_2 - \boldsymbol{x}_1||(\boldsymbol{n}^\top \boldsymbol{w}) \quad$ (where $\boldsymbol{n} = \frac{\boldsymbol{x}_2 - \boldsymbol{x}_1}{||\boldsymbol{x}_2 - \boldsymbol{x}_1||}$)
$\therefore y_2 - y_1 = 0 \iff \boldsymbol{n}^\top \boldsymbol{w} = 0$
$\therefore y_1 = y_2$ if and only if $\boldsymbol{w} \in \mathcal{H}_{\boldsymbol{n}}$.

3. We showed in 2 that $(\boldsymbol{X}\boldsymbol{w})_i$ and $(\boldsymbol{X}\boldsymbol{w})_j$ are equal if and only if $\boldsymbol{w} \in \mathcal{H}_{i,j}$, where $\mathcal{H}_{i,j} := \{\boldsymbol{w} : (\boldsymbol{X}_{j,:} - \boldsymbol{X}_{i,:})^\top \boldsymbol{w} = 0\}$ i.e. $\boldsymbol{w}$ is in the hyperplane normal to $\boldsymbol{X}_{j,:} - \boldsymbol{X}_{i,:}$, $\forall i, j \in \{1, ..., n\}, i \neq j$, i.e. when $\boldsymbol{w}$ is not in any of the $\binom{n}{2}$ hyperplanes. Since each hyperplane spans in a space which is one dimension lesser than the dimension of the vectors p, the $\binom{n}{2}$ hyperplanes cannot cover the full p-dimensional space. Hence, such a $\boldsymbol{w}$ exists (provided n is finite).

Question 2. Recall the variance of X is $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

1. Let X be a random variable with finite mean. Show $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
2. Let X and Z be random variables on the same probability space. Show that : $\mathrm{Var}(X) = \mathbb{E}_Z[\mathrm{Var}(X|Z)] + \mathrm{Var}_Z(\mathbb{E}[X|Z])$. (Hint : $\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}[X|Y]]$.)

Answer 2. 1. $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2)] = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

2. $\mathbb{E}_Z[\mathrm{Var}(X|Z)] + \mathrm{Var}_Z(\mathbb{E}[X|Z])$
$= \mathbb{E}_Z[\mathbb{E}[X^2|Z] - \mathbb{E}[X|Z]^2] + \mathbb{E}_Z[\mathbb{E}[X|Z]^2] - \mathbb{E}_Z[\mathbb{E}[X|Z]]^2$
$= \mathbb{E}_Z[\mathbb{E}[X^2|Z]] \underline{- \mathbb{E}_Z[\mathbb{E}[X|Z]^2]} \underline{+ \mathbb{E}_Z[\mathbb{E}[X|Z]^2]} - \mathbb{E}_Z[\mathbb{E}[X|Z]]^2$
$= \mathbb{E}_Z[\mathbb{E}[X^2|Z]] - \mathbb{E}_Z[\mathbb{E}[X|Z]]^2$
$= \mathbb{E}[X^2] - E[X]^2 = \mathrm{Var}(X)$

Question 3. Let $X \in \mathcal{X}$ be a random variable with density function f_X, and $g : \mathcal{X} \to \mathcal{Y}$ be continuously differentiable, where $\mathcal{X}$ and $\mathcal{Y}$ are subsets of $\mathbb{R}$. Let $Y := g(X)$, which is continuously distributed with density function f_Y.

1. Show that if g is monotonic, $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$.
2. Let $f_X(x) = \mathbf{1}_{x \in [0,1]}(x)$ and $f_Y(y) = \mathbf{1}_{y \in [0,2]}(y) \cdot \frac{y}{2}$. Find a monotonic mapping g that translates f_X and f_Y.

*3. Let $N_Y = \{y \in \mathcal{Y} : g(x) = y,\, g(x)' = 0 \text{ for some } x \in \mathcal{X}\}$. Show that in general if $g'(x) = 0$ at most finitely many times, for $y \in \mathcal{Y} \setminus N_Y$,

$$f_Y(y) = \sum_{x \in \{x : g(x) = y\}} \frac{f_X(x)}{|g'(x)|}$$

4. Let $X \sim \mathcal{N}(0,1)$, i.e. $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$, and $g(x) = x^2$. Determine $f_Y(y)$.

Answer 3.

Question 4. Let Q and P be univariate normal distributions with mean and variance μ, σ^2 and m, s^2, respectively. Derive the entropy $H(Q)$, the cross-entropy $H(Q, P)$, and the KL divergence $D_{\mathrm{KL}}(Q||P)$.

Answer 4. Entropy :

$$\begin{aligned}
H(Q) &= -\int q(x) \ln q(x) dx \\
&= -\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \ln\Big(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)\Big) dx \\
&= -\Bigg(\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \ln(\frac{1}{\sqrt{2\pi\sigma^2}}) dx + \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)(-\frac{(x-\mu)^2}{2\sigma^2}) dx\Bigg) \\
&= -\Bigg(\ln(\frac{1}{\sqrt{2\pi\sigma^2}}) \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx - \frac{1}{2\sigma^2} \int (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx\Bigg) \\
&= \ln(\sqrt{2\pi\sigma^2}).1 + \frac{1}{2\sigma^2}.\sigma^2 \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2} \\
&= \frac{1}{2}\ln(2\pi e \sigma^2)
\end{aligned}$$

Cross-Entropy :

$$
\begin{aligned}
H(Q,P) &= -\int q(x)\ln p(x)dx \\
&= -\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \ln\Bigg(\frac{1}{\sqrt{2\pi s^2}} \exp\Big(-\frac{(x-m)^2}{2s^2}\Big)\Bigg) dx \\
&= -\Bigg(\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \ln(\frac{1}{\sqrt{2\pi s^2}}) dx + \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) (-\frac{(x-m)^2}{2s^2}) dx\Bigg) \\
&= -\Bigg(\ln(\frac{1}{\sqrt{2\pi s^2}}) \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx - \frac{1}{2s^2}\int (x-m)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx\Bigg) \\
&= \ln(\sqrt{2\pi s^2}).1 + \frac{1}{2s^2}\int ((x-\mu)+(\mu-m))^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx \\
&= \frac{1}{2}\ln(2\pi s^2) + \frac{1}{2s^2}\Bigg(\int (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx \\
&\quad +2(\mu-m)\int (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx + (\mu-m)^2 \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) dx\Bigg) \\
&= \frac{1}{2}\ln(2\pi s^2) + \frac{1}{2s^2}\Bigg(\sigma^2 + 2(\mu-m).0 + (\mu-m)^2.1\Bigg) \\
&= \frac{1}{2}\ln(2\pi s^2) + \frac{\sigma^2+(\mu-m)^2}{2s^2}
\end{aligned}
$$

KL Divergence :

$$
\begin{aligned}
D_{\mathrm{KL}}(Q||P) &= -\int q(x)\ln\frac{p(x)}{q(x)} dx \\
&= -\int q(x)\ln p(x)dx + \int q(x)\ln q(x)dx \\
&= H(Q,P) - H(Q) \\
&= \frac{1}{2}\ln(2\pi s^2) + \frac{\sigma^2+(\mu-m)^2}{2s^2} - \frac{1}{2}\ln(2\pi e\sigma^2) \\
&= \underline{\frac{1}{2}\ln(2\pi s^2)} + \frac{\sigma^2+(\mu-m)^2}{2s^2} - \underline{\frac{1}{2}\ln(2\pi\sigma^2)} - \frac{1}{2} \\
&= \ln\Big(\frac{s}{\sigma}\Big) + \frac{\sigma^2+(\mu-m)^2}{2s^2} - \frac{1}{2}
\end{aligned}
$$